Second part of *Objective prior distribution of climate sensitivity, or... Koku and Toku looking for the shape of ignorance*

Prologue

Some time ago I wrote a comic version of my paper Pueyo (2012, Solution to the paradox of climate sensitivity, *Climatic Change* 113: 163-179), in which I sought the non-informative prior distribution of climate sensitivity. The comic featured a dialogue between two aliens named Koku and Toku. I presented it as a poster in the 2012 AGU Fall Meeting:

http://fallmeeting.agu.org/2012/eposters/eposter/gc43e-1086/

I realized that there was much confusion between my method and Nicholas Lewis', who also claims to be using the non-informative prior distribution. One of the main differences is that my method follows Edwin T. Jaynes' criterion (Jaynes is best known for having introduced the maximum entropy principle), while Lewis (like Jewson et al.) follows Harold Jeffreys' criterion. I do not mean that Jeffreys' criterion is useless, but I do not think that it is the criterion of choice in our context.

I had the fortune of "hearing" a second conversation between Koku and Toku, in which they compare these two options. I reproduce this second dialogue below, in two parts, 2A and 2B. Part 2A can be read alone. Part 2B is for readers who want more.

Dialogue between Koku and Toku, Part 2A

Toku: An asteroid is about to hit Loku's house. I have to talk to him immediately. He went on vacation to the Earth's surface, but I know nothing else. Which prior distribution should I use to find him?

Koku: If we have no idea where he is, all positions are equally likely. Therefore, the non-informative prior distribution is uniform all over that planet's surface.

T: This is indeed what my intuition would say, but can you justify it rationally?

K: It's simple. The non-informative distribution of a position cannot be, for example, Gaussian with mean zero, because you can put the *zero* value (the origin of the coordinate system) wherever you want, and each choice will give you a different map of probabilities. The uniform is the only distribution that is independent of the origin that you choose. (Note the contrast with climate sensitivity S, where there is no arbitrariness in the zero value of S; there is, however, arbitrariness in the zero value of $\log(S)$, which depends on the units chosen to measure S, so the non-informative prior of S has to be uniform on a logarithmic scale).

T: Is this the kinda logic that, you said, earthlings were not using before Jaynes introduced it to his planet (and most do not use yet)?

K: Right, it's what Jaynes named the *invariant groups* criterion.

T: You seem very convinced of this criterion, but I heard that, among objective Bayesian earthlings, it's more usual to abide by a different criterion, due to Jeffreys. What prior should I use if I decide to follow this other criterion?

K: Apart from discussing about priors, what else are you planning to do to find Loku?

T: Oh I see, you are angry because I wanted to know about criteria other than yours, and now you are trying to deviate the conversation...

K: No problem, I'll answer your question, but first, please, answer mine.

T: Your reactions are so weird sometimes... Well ok, I'm going to install a locator on top of the Eiffel Tower.

K: So you could not think of any more discreet place... and you say I'm weird...

T: I don't know many other places in that planet...

K: Well, and what is the error of this locator like?

T: The error is approximately Gaussian. Another property is that the larger the distance, the larger the error.

K: Then, the closer you are to Eiffel Tower, the larger is the prior probability according to Jeffreys' criterion. This means that Loku is most likely to be in Paris.

T: But Loku does not know that I am planning to install a locator in Paris. So, why should he have gone to Paris?

K: Well, I do not really think that there is any reason for you to expect finding him in Paris. I still think that, before your locator begins to give data, you should consider that all positions are equally likely. Whatever we can tell a priori about the position of Loku, it cannot depend on the kind of data that you are planning to obtain in the future. This is why I don't use Jeffreys' prior.

Dialogue of Koku and Toku, Part 2B

T: So if Jeffreys prior is so absurd, why are there earthlings using it?

K: I would distinguish two main groups. Some are scientists too busy in their own fields to navigate the intricacies of the many schools of statistics, who have been told that Jeffreys' prior is non-informative and have assumed that this is correct. Others, including the best known successors of Jeffreys (such as Bernardo and Berger) are aware that what this author found wasn't a non-informative prior, but they do not think that there is any way to find a true non-informative prior. As an alternative, they use Jeffreys' or closely related priors for two reasons. First, they consider that it is good to have a standard of reference to be used by all scientists in order to make their results comparable, like the standards that earthlings have for their units of measurement (an argument put by the own Jeffreys in his later works). For this reason, they do not talk of *non-informative priors*, but of *reference priors*. Second, they consider that Jeffreys' or some related prior are good candidates to be elected as reference priors, because, when combined with the kind of observations that you are going to use, the resulting posterior will have good statistical properties. Given enough data, they will converge to the right value without problems such as major artifacts created by the prior. However, if you do not have so many data, your posterior distribution will display some features reflecting nothing else than a convention. Of course, apart from these groups, there might be other earthlings using it for other reasons.

T: The discussions between Jeffreys and Jaynes must have been strong...

K: Not at all. Jaynes took an early approach by Jeffreys as his starting point, and, as far as I can tell,

he never stressed the differences between their approaches. However, from this common starting point, they reached conclusions whose differences are really important, in spite of being terribly subtle. Jaynes' invariant groups method can be applied without reliance on the experimental design (e.g. whether your locator is on Eiffel's Tower or on Taj Mahal). Therefore, if, as I think, non-informative distributions exist, you can get them from Jaynes but not from Jeffreys. However, Jaynes' method is or appears to be much trickier than Jeffreys, because it requires thinking about the nature of the problem, while Jeffreys' consists of applying a formula. This has probably helped Jeffreys' method being more popular than Jaynes'...

T: I see. If I have many data (or my detector has very little error), both Jaynes and Jeffreys will give me almost the same result, but Jeffreys' path will be easier and safer because I won't have to think about the nature of the problem. If I have little data, and non-informative distributions really exist, only Jaynes can give me the correct answer. If I have little data, and non-informative distributions do not exist, Jeffreys' approach will be again easier and safer than Jaynes, but, otherwise, its results will be neither more nor less realistic than those obtained from Jaynes or many other possible priors... In this case, the answer will just be that there is no well-defined probability distribution. I could still choose Jeffreys in order to contribute to the establishment of a convention, but this convention will not help me to find Loku (or will not help earthlings in their climate policy).

K: You understood perfectly everything I was trying to say. However, there is something else. Jeffreys' prior is not really as easy and safe as it seems to be. You have to be very careful when defining the likelihood function. Some subtle features, which have little relevance for frequentist statistics, can have dramatic consequences when used to derive a prior by Jeffreys' method. For example, when applying this method, there is much risk in common simplifications such as assuming that all noises are Gaussian.

T: And is it so important to wonder about the correct prior?

K: Well, look at the cost-benefit relation. The benefit of using a good instead of a bad prior is not necessarily less than the benefit of using a high-quality locator instead of an average one. However, a better locator will cost you much extra money (assuming you get it in a monetized planet), while a better prior will cost you *only* extra thinking.

T: I see. So your point of view is that it is really worth trying to use the correct prior, and that the correct prior will result from applying the invariant groups method. Are you sure of this? Remember that it is serious, that Loku might lose his house if we do not find him...

K: I am never sure of anything. Provisionally, however, the invariant groups approach is the only one that I find really coherent. However, you are not obliged to trust me. Being an issue in which *experts* disagree and do so for deep philosophical reasons, you cannot rely on *expert opinion* alone. If you want to walk on solid grounds, the only think you can do is to spend some time comparing the different rationales, and, then, to decide according to your own logic and intuition.

T: Thanks, Koku, this was really interesting. The only problem is that, according to my calculations, the asteroid must have already crashed into Loku's house a few seconds ago...

SALVADOR PUEYO